METHODS

The CanMap sample collection. Our study is part of the CanMap project (A.R.B. *et al.*, manuscript submitted), which isolated genomic DNA from blood samples collected from domestic dogs (*C. familiaris*, $n = 912$) and from tissue and blood samples from grey wolves (*C. lupus*, $n = 225$), coyotes (*C. latrans*, $n = 60$), putative dog–wolf hybrids ($n = 17$), red wolves (*C. rufus*, $n = 12$), Mexican wolves (*C. l. baileyi*, $n = 10$), Ethiopian wolves (*C. simensis*, $n = 4$), black-backed jackals (*C. mesomelas*, $n = 6$), golden jackals (*C. aureus*, $n = 2$), and a side-striped jackal (*C. adustus*, $n = 1$; Supplementary Table 1). Domestic dog samples were obtained through American Kennel Club (AKC) sanctioned dog shows, speciality events, breed clubs, and veterinary clinics. Three-to-twelve dogs from each breed from each of 81 AKC-recognized breeds and four semi-domestic lineages (Africanis, Canaan dog, dingo, and New Guinea singing dog) were used in the analysis. Specifically, the semi-domestic dingo and New Guinea singing dog are ancient breeds that were probably established more than 4,000 years ago and have existed in isolation from wolves[32].

The Affymetrix Canine Mapping Array version 2 contains SNPs that were ascertained by aligning sequence reads to the boxer genome assembly (CamFam2). A large number of SNPs were discovered as heterozygous sites in the boxer genome (here denoted as boxer × boxer SNPs), and further SNPs were found by aligning sequence reads from other breeds or wild canids to the boxer genome. These extra SNPs can be categorized by the sequence aligned to the boxer as follows: (1) the standard poodle (CanFam1); (2) one of nine dog breeds; (3) one of four wolf populations (Alaskan, Chinese, Indian or Spanish wolves); and (4) coyote (see Supplementary Table 7).

Breed groupings. Several analyses are based on specific dog breed groupings for comparison purposes (Supplementary Table 1). We define ancient breeds as those that are divergent genetically (Fig. 1), corroborated by previous genetic studies[33] and, in most cases, are known to have originated in ancient cultures more than 500 years ago[8,9]. Furthermore, we used previously defined geographical breed groupings[3] (Africa, America, east Asia, Europe, Siberia, southeast Asia and southwest Asia breed groups) and functional and phenotypic breed groupings in common usage by dog breeders[8,9,33] (ancient, spitz, toy, spaniels, scent hounds, working dogs, mastiff-like breeds, small terriers, retrievers, herding and sight hounds).

Identification of recently related individuals in the sample. We used PLINK[34] to obtain pairwise estimates of identity by state (IBS). From the Yellowstone National Park wolves in the data set ($n = 19$), known pedigree relationships were used to calibrate IBS scores[35]. A minimum score of IBS > 0.8 indicated a relatedness status of half-siblings, and values below this level were used to identify a set of unrelated wild canids.

Single-SNP measures of genetic diversity. Single-marker descriptive statistics (observed/expected heterozygosity and polymorphism) were estimated using PLINK[34] for the complete SNP data set (Fig. 2). Observed heterozygosity was also estimated using only SNPs ascertained from the grey wolf and boxer sequence comparisons (Supplementary Fig. 12). We used microsatellite genotype data from a previous study[4] for an independent comparison of observed heterozygosity from loci with different mutational properties and ascertainment schemes (Fig. 2 and Supplementary Fig. 13).

STRUCTURE analysis. We used the Bayesian inference program STRUCTURE[36] to assess genetic partitions and admixture for the 43,953-SNP data set (linkage disequilibrium (LD) pruned, $r^2 < 0.5$). We used 5,000 burn-in iterations and 15,000 Markov chain Monte Carlo (MCMC) iterations in STRUCTURE, with three repetitions of these parameter settings. The alpha and likelihood statistics were verified to reach convergence before 5,000 burn-in and 15,000 MCMC iterations were completed during each repetition for each number of assumed populations analysed. We analysed domestic dogs and Old World wolves to resolve the ancestry of domestic dogs; hence, we included only one dog per breed for the analysis ($n = 85$; Supplementary Fig. 5). We also included only Old World wolf populations because they may be closely related to the direct ancestors of domestic dogs and we used unrelated individuals from IBS estimates. We sampled China ($n = 9$), central Asia ($n = 3$), the Middle East ($n = 7$) and Europe ($n = 43$). We excluded wolves from highly inbred populations (Italy, Spain and Sweden)[24] to avoid their early partitioning in the analysis. No dog–wolf hybrids were found in the full sample of modern breeds ($n = 801$) as determined with the program smartpca in the Eigensoft package[37].

From the dog–wolf PCA (see Supplementary Note B) we identified 20 SNPs with the highest loadings on PC1 as input for an additional STRUCTURE analysis to determine the posterior probability of assignment for dogs and wolves to their corresponding species (Supplementary Table 2). Results were plotted using the circular visualization program CIRCOS (http://mkweb.bcgsc.ca/circos/; Supplementary Fig. 5).

After the initial partitioning of modern domestic dogs from wild canids for $K = 2$ (Supplementary Fig. 5), ancient breeds are separated from other canids when a third population is assumed ($K = 3$). Our results show uniquely that Canaan dog, dingo, New Guinea singing dog, and Alaskan Eskimo dog are members of this cluster of ancient breeds, and confirm previous results showing basenji, Afghan hound, samoyed, Saluki, Canaan dog, chow chow, Chinese Shar Pei, Akita, Alaska malamute, and Siberian husky belong in that group[3,4,8,9].

Analysis of molecular variance. The IBS matrix was put into ARLEQUIN v3 to analyse molecular variance with 10,098 permutations for significance testing[38]. We defined three different analysis groups (see Fig. 1 and Supplementary Table 5): (1) breed groups in Fig. 1; (2) geographical dog breed groups (Supplementary Table 1); and (3) wolves and dogs as separate groups.

32. Savolainen, P., Leitner, T., Wilton, A., Matisoo-Smith, E. & Lundeberg, J. A detailed picture of the origin of the Australian dingo, obtained from the study of mitochondrial DNA. *Proc. Natl Acad. Sci. USA* **101,** 12387–12390 (2004).
33. Sablin, M. & Khlopachev, G. The earliest Ice Age dogs: evidence from Eliseevichi I. *Curr. Anthropol.* **43,** 795–799 (2002).
34. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
35. vonHoldt, B. *et al.* The genealogy and genetic viability of reintroduced Yellowstone grey wolves. *Mol. Ecol.* **17,** 252–274 (2008).
36. Pritchard, J., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).
37. Price, A. *et al.* Principle components analysis corrects for stratification in genome-wide association studies. *Nature Genet.* **38,** 904–909 (2006).
38. Schneider, S., Roessli, D. & Excoffier, L. Arlequin: a software for population genetics data analysis v.2.000 (Genetics and Biometry Lab, Department of Anthropology, University of Geneva, 2000).

©2010 Macmillan Publishers Limited. All rights reserved